A New Approach to Train LSTMs, GRUs, and Other Similar Networks for Data Generation

Ravin Kumar(✉)

Department of Computer Science,
Meerut Institute of Engineering and Technology, Meerut 250005,
Uttar Pradesh, India
ravin.kumar.cs.2013@miet.ac.in

Abstract. Deep Learning has played a major role in the field of Artificial Intelligence in solving some of the toughest problems. In this paper, a training approach for deep learning model based on LSTMs, GRUs, and other similar variants is proposed. The effectiveness of this approach is checked on two different models, the first model was based on LSTMs and the second model was based on GRUs. To maintain fairness during comparison few parameters were made constant and then different tests were carried out by varying the dropout. It was found that both LSTMs and GRUs models which were trained using the proposed approach had quickly reduced their training loss without underfitting or overfitting, on the data and converges much faster than the traditional approach. A comparative study is done on a text generation task to show the differences between the quality of data generated by our proposed model with respect to the traditional model.

Keywords: Text generation · LSTM · GRU · Recurrent neural network

1 Introduction

Artificial Intelligence is playing a crucial role in solving problems which were tough to solve using traditional programming approaches. It is effective in tasks such as image classification, text generation, speech recognition etc. Deep learning methods have been available since the late 90s though their applications have been emerging after development of graphic cards and high storage capacities that reduce the training time from a year to weeks on humongous amount of data.

The objective of training a model is to minimize loss during the learning process without over fitting or under fitting on the dataset. Conventional Neural Networks required carefully engineered features to be fed into the network. This has been resolved with the help of deep learning where the huge network automatically figures out the relevant features given enough data. The innovation of various new layers has led to new state of the art techniques which surpass the leading results. This paper proposes a new approach in training Deep Learning Models including LSTMs, GRUs, and other similar variants.

© Springer Nature Singapore Pte Ltd. 2019
M. Prateek et al. (Eds.): NGCT 2018, CCIS 922, pp. 163–171, 2019.
https://doi.org/10.1007/978-981-15-1718-1_14

2 Related Work

Hochreiter et al. [1] proposed LSTMs as a new approach for traditional recurrent neural networks as a solution to their vanishing gradient problem. Mikolov et al. [2] came up with an effective approach to represent words in the form of embeddings, such that similar words are closer to each other in their vector space. Chung et al. [3] suggested GRUs as an alternative approach for LSTMs having comparable results. Graves et al. [4] developed Bidirectional LSTMs, which were better than the traditional unidirectional LSTM. Srivastava et al. [5] proposed Dropout mechanism to prevent over fitting in Neural networks, which was later used by Cheng et al. [6] in LSTMs to optimize its performance. All these advances lead to the development of many End-to-End Deep Learning models, which required less preprocessing. Graves et al. [7] used Deep LSTM Recurrent Neural Networks for Speech Recognition on the TIMIT phoneme recognition. Palangi et al. [8] used LSTMs for generating deep sentence embeddings. Nayebi et al. [9] compared different RNNs for algorithmic music generation and found that result obtained from LSTMs were better than GRUs. Wen et al. [10] proposed a statistical language generator based on semantically controlled LSTMs, and it received more scores on informativeness and naturalness than other available models. Xiao et al. [11] suggested a system for music generation using LSTMs, and their obtained results were similar to input original music. Huang et al. [12] build a generative model using two layers stacked LSTMs RNN architecture and were able to produce music comparable to those obtained from sophisticated time series models. Sutskever et al. [13] Sequence to Sequence models used LSTMs as its basic unit, and were better at tasks related to sequence mapping, and one of the very few models available for End to End Translation Systems.

3 Proposed Training Approach

Bidirectional LSTMs and bidirectional GRUs represent data in much efficient manner than unidirectional LSTMs and GRUs respectively. But in the case of data generation or other similar task like music generation, we can not use Bidirectional RNNs for data generation purposes as it requires input data in both forward and reverse order which is not possible to provide at inference time.

The proposed model is also a solution to this problem, as it requires data in forward and reverse order of input sequence at training time, while at the time of inference it takes data in forward order. This enables our training approach to have better performance than unidirectional LSTMs and GRUs based models in data generation task. The proposed approach uses an additional network to train the model, we called this additional network ‘Shadow Network’. It is the replica of traditional model which was supposed to generate data, except it take inputs in the reverse order of the input sequence. Output produced by the Shadow Network append with the output of traditional model at each instance and is send to a feed forward neural network. This feed forward neural network help the formed combined model to learn the sequence in non-reverse order. At each time step, the output from traditional model LSTM/GRU cell is

also send to another feed forward neural network whose main purpose is to make traditional model learn the original input sequence.

To get visual understanding of proposed mechanism let us assume a traditional LSTM or GRU based model M_1 designed for data generation task as represented in Fig. 1. NN represents a feed forward neural network which is used to train model on input sequence and, later used for generating the sequence on inference phase.

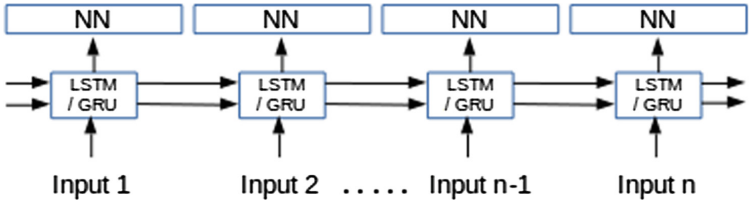

Fig. 1. LSTM/GRU based traditional model M_1 during training time.

When proposed 'Shadow Network' is attached to traditional model M_1, the overall architecture of the model get changed as represented in Fig. 2.

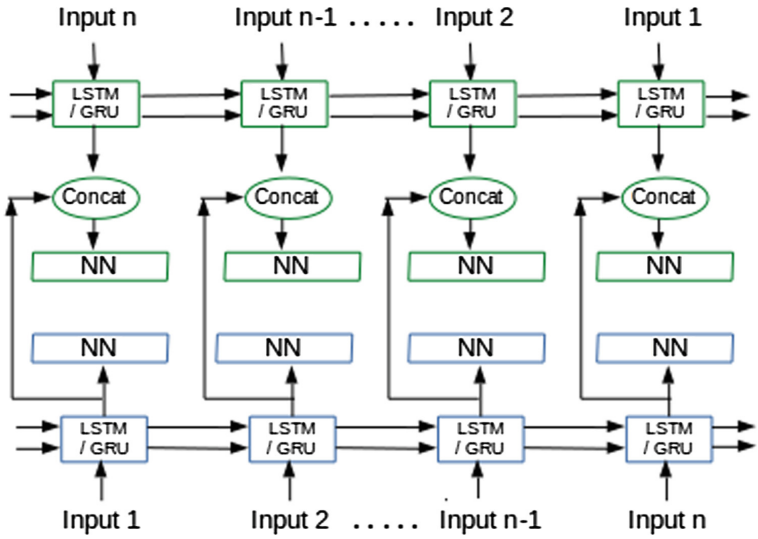

Fig. 2. Upper network (in green) representing the Shadow network attached to feed forward neural network, and the lower network (in blue) representing the model M_1 attached to feed forward neural network. (Color figure online)

During inference phase Shadow Network is removed and model M_1 is used to generate an output sequence as shown in Fig. 3.

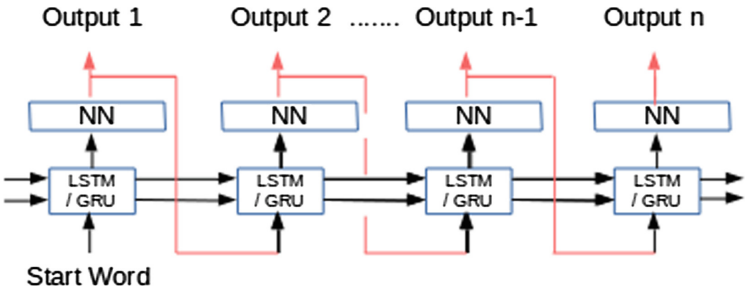

Fig. 3. Model M_1 predicting output sequence during inference phase.

4 Result and Discussion

We performed variety of experiments to analyze the performance of models trained using proposed network vs models trained without it. Separate models were designed for LSTM and GRU cells to evaluate their performance using the proposed training approach. Different test cases were designed by varying the dropout value for the models. During experiment some parameters were made constant in order to get an unbiased comparison report. We have also made our experiment data available at our github repository [14]. Given below are the value of parameters used in the experiment (Table 1).

Table 1. List of parameters used for training.

S.no	Parameter	Value	Details
1	Dataset	Simpson dataset [14]	Textual data
2	Epoch	200	Constant value
3	Batch size	250	Constant value
4	Rnn size	512	Constant value
5	Sequence length	21	Constant value
6	Learning rate	0.001	Variate during training
7	LSTM/GRU Layers	3	Constant value
8	Dropout	0%, 25%, 50%	Different value for different test cases (but remains constant in each test case)

When proposed approach is used to train traditional LSTM model M_1 then the training loss of traditional LSTM model reduced much faster than that of a simple LSTM based RNN model because our shadow network enabled the analysis of data in both forward and reversed order during the training time, which helps the deep learning model understand the data in an efficient manner. While during inference time shadow network became inactive and we use only the traditional model M_1 to generate data.

Dropout generally helps deep learning model to better train on the data. But even after applying various value of dropout our approach performed better than the traditional approach (Figs 4, 5 and 6).

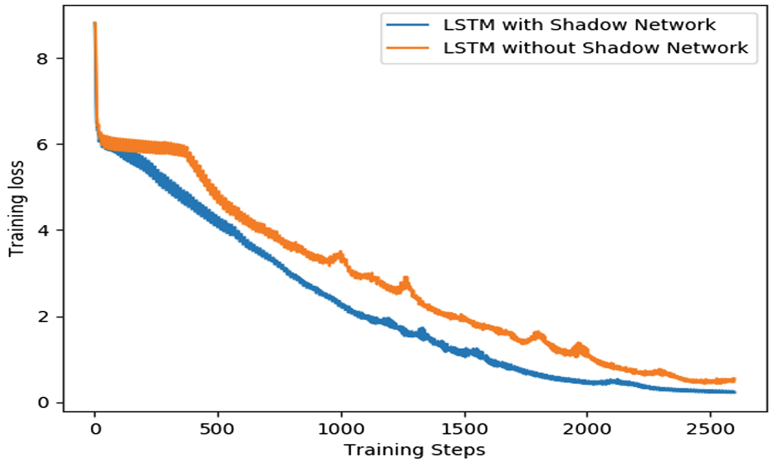

Fig. 4. Comparative study over LSTMs when dropout is kept at 0 percent.

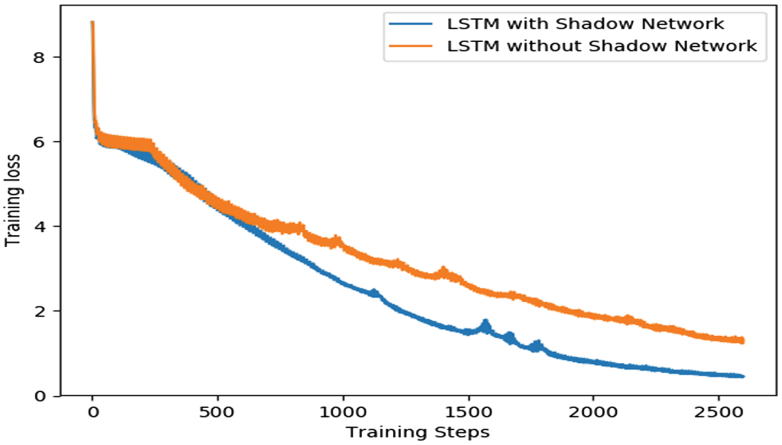

Fig. 5. Comparative study over LSTMs when dropout is kept at 25 percent.

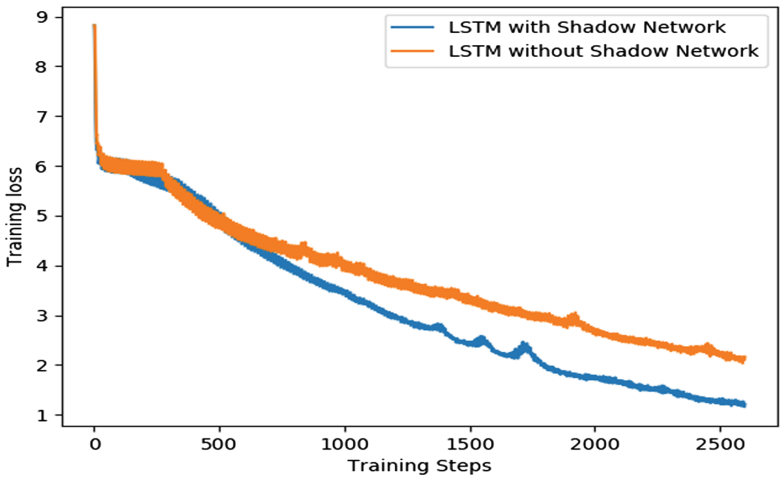

Fig. 6. Comparative study over LSTMs when dropout is kept at 50 percent.

When GRU based recurrent neural networks are examined for comparative study of training loss between the traditional and our proposed approach the results obtained after varying dropout value also support our proposed approach (Figs 7, 8 and 9).

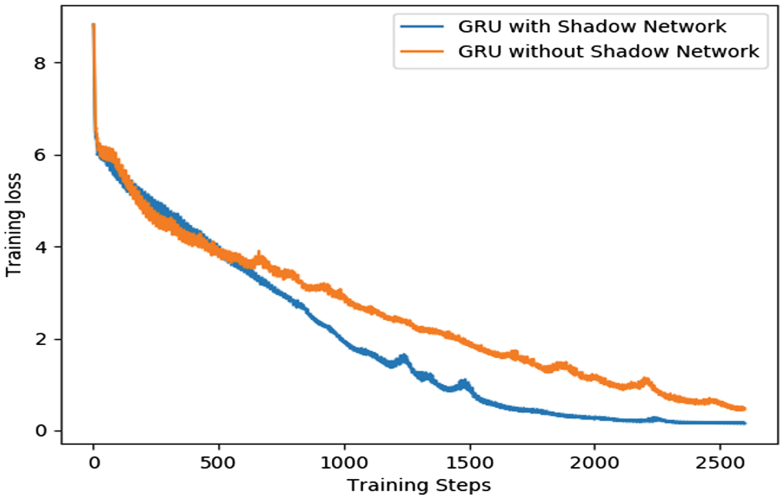

Fig. 7. Comparative study over GRUs when dropout is kept at 0 percent.

Various mechanisms exist to evaluate the quality of data generated in machine translation but to perform a comparison on the quality of data generated in a simple data generation task is very tough problem, as no specific mathematical mechanism is known which can perform a comparative analysis on quality of the data generated between two recurrent neural network based architectures with respect to training data.

We have used human assistance to evaluate the quality of data generated by our proposed mechanism during the inference period. For the evaluation of our proposed model we distributed the data generated during inference by both traditional and our proposed approach to people and asked them to rank each sentence from 1 to 5 numbers, where 1 represents least quality data and 5 represent high quality data. After getting the evaluation score on data generated during inference charts are drawn to get visual comparison of the quality of data produced by using the proposed approach (Figs 10 and 11).

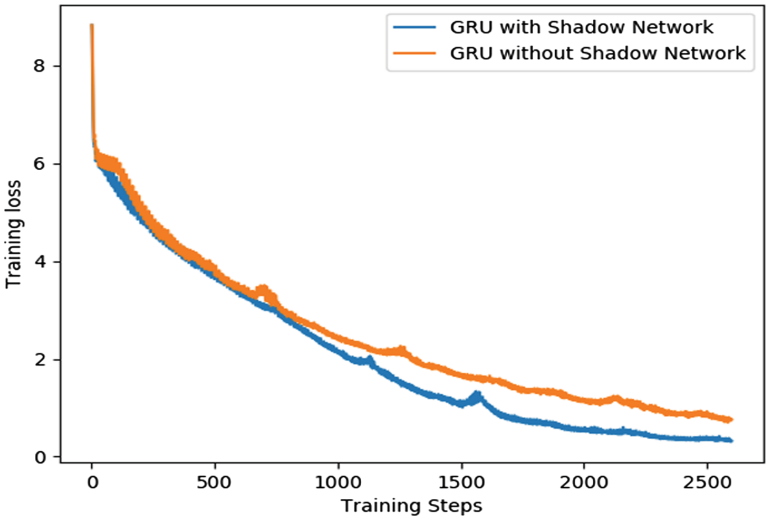

Fig. 8. Comparative study over GRUs when dropout is kept at 25 percent.

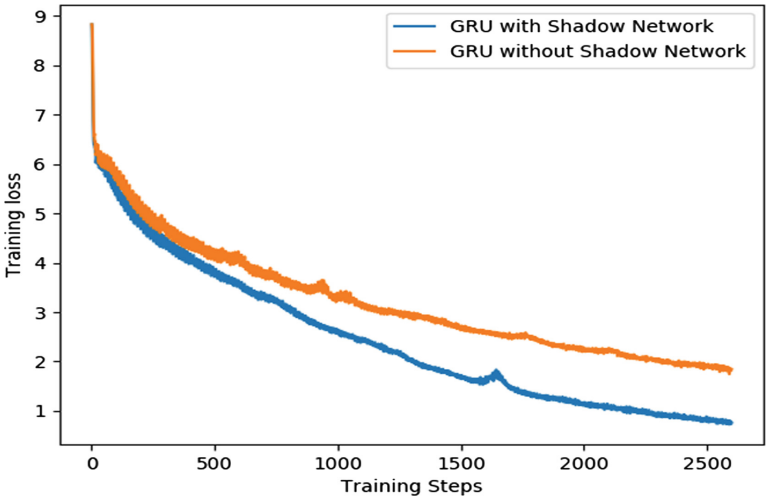

Fig. 9. Comparative study over GRUs when dropout is kept at 50 percent.

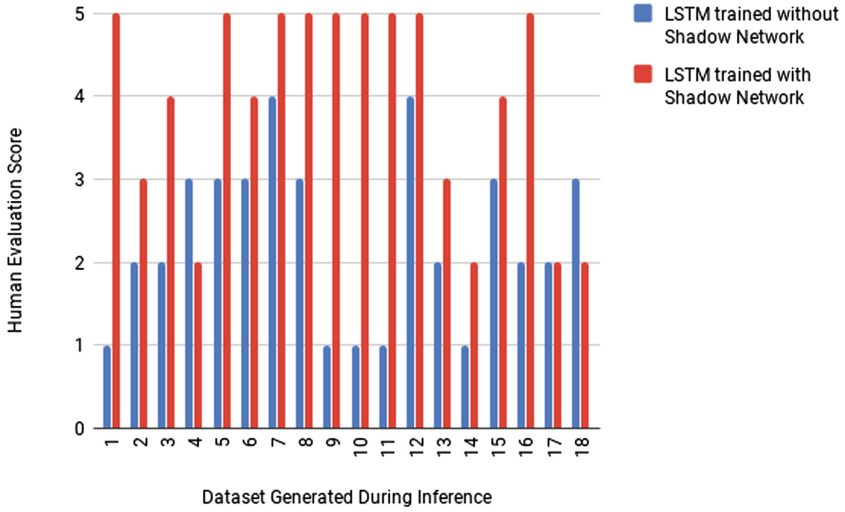
LSTM trained without Shadow Network
LSTM trained with Shadow Network
Human Evaluation Score
Dataset Generated During Inference

Fig. 10. Comparative study on quality of data generated by LSTM based architectures.

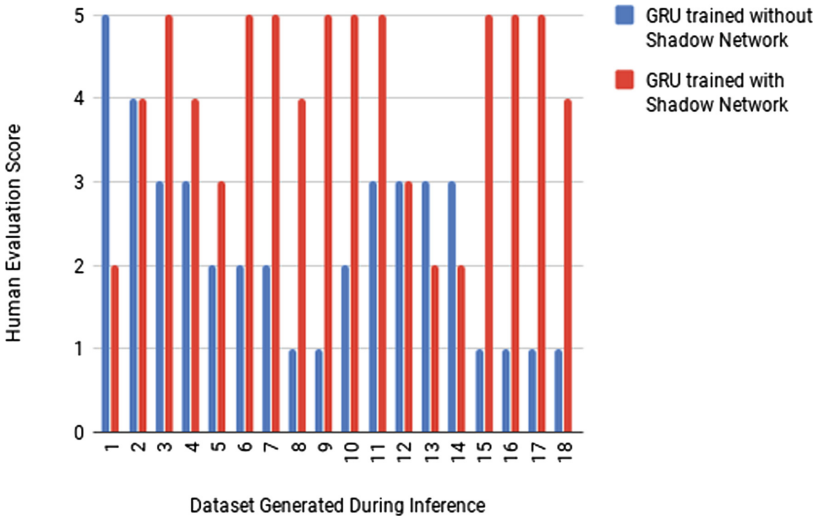
GRU trained without Shadow Network
GRU trained with Shadow Network
Human Evaluation Score
Dataset Generated During Inference

Fig. 11. Comparative study on quality of data generated by GRU based architectures.

These results suggest that shadow network helps improving the quality of data generated at inference when compared with model trained using traditional approach.

We have also designed a twitter bot [15] which regularly posts self generated quotes on twitter. This bot is trained using our proposed approach over quotes data which we collected from the web. It was deployed with the aim of providing a simple application of our research work with an insight on quality of data produced during the inference phase.

5 Conclusion

In this paper, a new and effective training approach is proposed for deep learning models based on LSTM, GRU and other variants for data generation. The proposed approach reduces the training loss under similar training parameters leading to better performance. Our approach opened a new door for performing various research works on data generation related models, specially the Decoder part of Sequence-to-Sequence model with or without having attention or other related mechanisms.

References

1. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
2. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space, Cornell University Library (2013). arXiv:1301.3781
3. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, Cornell University Library (2014). arXiv:1412.3555
4. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)
5. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. **15**, 1929–1958 (2015)
6. Cheng, G., Peddinti, V., Povery, D., Manohar, V., Khudanpur, S., Yan, Y.: An exploration of dropout with LSTMs. In: INTERSPEECH, Annual Conference of the International Speech Communication Association, Stockholm, Sweden, 20 August 2017
7. Graves, A., Mohamed, A.-R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649 (2013)
8. Palangu, H., et al.: Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval, Cornell University Library (2015). arXiv:1501.06922v3
9. Nayebi, A., Vitelli, M.: GRUV: Algorithmic Music Generation using Recurrent Neural Networks (2015). https://cs224d.stanford.edu/reports/NayebiAran.pdf
10. Wen, T., Gasic, M., Mrksic, N., Su, P.-H., Vandyke, D., Young, S.: Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In: Proceedings of EMNLP, Lisbon, Portugal, pp. 1711–1721, September 2015
11. Xiao, Y., Xiong, S., Duan, P.: Music generation system based on LSTM. In: 4th International Conference on Electrical & Electronics Engineering and Computer Science, Advances in Computer Science and Research, vol. 50. ICEEECS (2016)
12. Huang, A., Wu, R.: Deep Learning for Music, Cornell University Library (2016). arXiv:1606.04930
13. Sutskever, I. Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Proceedings of Advances in Neural Information Processing Systems, vol. 27, pp. 3104–3112 (2014)
14. Github repository. https://github.com/mr-ravin/ShadowNetwork. Accessed 3 Nov 2018
15. Heilige Quotes bot on Twitter. https://twitter.com/HeiligeQuotes. Accessed 3 Nov 2018